Confidence Intervals: Evaluating and Facilitating Their Use in Health Education Research

Jing Zhang, Bruce W. Hanik, and Beth H. Chaney

Abstract

Health education researchers have called for research articles in health education to adhere to the recommendations of American Psychological Association and the American Medical Association regarding the reporting and use of effect sizes and confidence intervals (CIs). This article expands on the recommendations by (a) providing an overview of CIs, (b) evaluating the use and interpretation of CIs in selected journals in health education, (c) presenting how to calculate CIs using statistical software, and (d) suggesting how to interpret and use CIs. Thirty-three articles in the *American Journal of Health Behavior* and *Health Education & Behavior* were evaluated. The evaluation showed that although CIs were reported in approximately half of the evaluated quantitative studies, they were not interpreted in any of the studies. The lack of interpretation of CIs indicates that health educators might not fully understand the meaning of CIs and consequently could not make use of CIs except for presenting the numbers. This article intends to increase health researchers' understanding of CIs, encourage the practice of thinking meta-analytically, and facilitate the use of CIs in the future.

The call for health educators to adhere to the American Psychological Association's (APA, 2001) and the American Medical Association's (AMA, 1998) requests regarding the reporting of effect sizes and confidence intervals (CIs) in research reports and articles is becoming more apparent in the health education literature. The latest *Publication Manual* of the APA highly recommended the use of CIs in research articles (APA, 2001). The *Publication Manual* regarded CIs as "in general, the best reporting strategy" (APA, 2001, p. 22). Similarly, the AMA *Manual of Style* (1998) indicates that reportage of CIs is preferred over *p* values, because they "convey information about precision as well as statistical significance" (p. 539). Additionally, studies conducted by Watkins, Rivers, Rowell, Green, and Rivers (2006), Rivers and Rowell (in press), and Buhi (2005) strongly encourage increased use and reporting of effect sizes and CIs for effect size calculations. Assuming these recommendations made by the APA *Publication Manual*, AMA *Manual of Style*, and the cited health education researchers will lead to better accumulation and application of the scientific knowledge, the field of health education could benefit from having its journals follow these recommendations. The purpose of this article is to expand on these recommendations by (a) providing an overview of CIs, (b) evaluating the use and interpretation of CIs in selected journals in health education, (c) presenting how to calculate CIs using statistical software, and (d) suggesting how to interpret and use CIs. The intended results of this article are to increase health researchers' understanding of CIs, provide a snapshot of the frequency and quality of CIs' use in health research, and facilitate the use of CIs by health researchers in the future.

An Overview of Confidence Intervals

Defining Confidence Intervals

A CI is an interval estimation of the population parameter (population characteristic). Computed with the sample statistic, a CI involves a range of numbers that possibly include the population parameter. A CI has four noteworthy characteristics. First, for a given sample size, at a given level of confidence, and using probability sampling, there can be infinitely many CIs for a particular population parameter. The point estimates and endpoints of these CIs vary due to sampling errors that occur each time a different sample is drawn (Thompson, 2002). Second, the CI reported by a certain study is just one of these infinitely many CIs. Third, the percentage of these CIs that contains the population parameter is the same with the level of confidence. Fourth, whether a certain CI reported by a study contains the population parameter is unknown. In other words, the level of confidence is applied to the infinitely many CIs, rather than a single CI reported by a single study (Thompson, 2006).

The following is an example to help illustrate the characteristics mentioned above. In a study investigating the predictors of current smoking among Vietnamese American men, Wiecha, Lee, and Hodgkins (1998) reported that higher educational level is negatively associated with current smoking (OR=0.8; 95% confidence interval 0.7 to

* Jing Zhang, MS, Doctoral Student; Department of Health and Kinesiology, Texas A&M University, College Station, TX 77843-4243; Telephone: 979-847-9587; Fax: 979-862-2672; E-mail: jingostar1980@hlkn.tamu.edu

Bruce W. Hanik, MS, Doctoral Student; Department of Health and Kinesiology, Texas A&M University, College Station, TX 77843-4243; Chapter: Alpha Pi.

Beth H. Chaney, PhD, CHES; Assistant Professor; Department of Health Education and Promotion, East Carolina University, 3205 Carol G. Belk Building, Greenville, NC 27858; Telephone: 252-328-1611; E-mail: chaneye@ecu.edu; Chapter: Beta Theta

* Corresponding author

0.9). The "95%" refers to the level of confidence ($1-\alpha$), which is the complement of the level of significance $\alpha=0.05$ (Hinkle, Wiersma, & Jurs, 2003). With the sample size of 774 and level of confidence of 95%, Wiecha et al. drew a probability sample and got an interval of 0.7 to 0.9. With the same sample size, level of confidence, and sampling method, another researcher might get a different OR and interval, which is OR=0.6, 95% confidence interval 0.3 to 0.9. The difference in point estimates and endpoints of the two CIs results from sampling error. If researchers keep drawing samples using Wiecha et al.'s procedures, they will have infinitely many intervals. Ninety-five percent of these intervals will contain the population parameter. However, whether Wiecha et al.'s or any other researcher's interval contains the population parameter is unknown.

Hinkle et al. (2003) explained the meaning of a 95% confidence interval of 2.20-2.70 as follows:

> Theoretically, suppose we compute the sample means of all possible samples of size 20 and constructed the 95-percent confidence intervals for the population mean using all these sample means. Then 95 percent of these intervals would contain μ [population parameter] and 5 percent would not. Note that we *cannot* say that the probability is .95 that the interval from 2.20 to 2.70 contains μ. Either the interval contains μ or it does not. (p. 205)

Computing Confidence Intervals

The CI for non-effect size statistics and the CI for effect sizes are computed differently. For non-effect size statistics, such as mean, a formula is used to calculate the CI. Hinkle et al. (2003) provided a general formula (p. 203): CI = Statistic ± (Critical value) (Standard error of the statistic). This formula shows that the standard error of the statistic determines the width of the CI. The standard error of the statistic refers to the standard deviation of the sampling distribution of the sample statistic. The larger the standard error, the wider the CI, and the less precise the interval estimate.

CIs for effect sizes cannot be computed with formulas. Instead, a statistical procedure (available in computer software such as SPSS)—iteration—must be performed to compute CIs for effect sizes (Thompson, 2006). Thompson (2006) noted, "As conventionally performed, iteration involves a process of initially guessing a solution, and then repetitively tweaking the guess until some statistical criterion is reached" (p. 207). Cumming and Finch (2001) and Kline (2004) have more detail on computation of CIs for effect sizes using iteration (Thompson, 2006).

The Importance of Confidence Intervals: Indicating Precision and Facilitating Meta-analytic Thinking

A CI displays the full range of hypothetical values of a parameter that cannot be rejected, thus is more informative than a statistical significance test (which only focuses on one null hypothesis value), although most of the information provided by a CI is not about statistical significance (Smithson, 2003). A CI also reveals the precision of the interval estimate—the narrower the width, the more precise the estimate. However, a CI tells nothing about whether it contains the parameter. Researchers might get excited about a 95% CI that does not subsume the null hypothesis parameter value, indicating that the statistic around which the CI is constructed is statistically significant. They might get even more excited when this CI is narrow, indicating that the CI is precise. Nevertheless, this narrow and "not subsuming null hypothesis parameter value" CI can still be among the 5% of CIs that does not contain the parameter (Thompson, 2006).

With this uncertainty, researchers may ask: Why are CIs important? CIs are important, not as isolated CIs reported by single studies, but as an addition to the collective body of all relevant CIs from previous studies. The most thoughtful use of CIs involves comparing CIs across studies to reveal the true parameter, regardless of whether the CIs subsume the null hypothesis parameter value, or whether the statistics around which CIs are constructed are statistically significant (Thompson, 2006). Academic journals' focus on statistical significance, rather than on documenting and integrating CIs, contributes to a publication bias where only statistically significant results are published, but non-significant results are not, creating an incomplete and biased picture in the literature (Thompson, 2002). The broader picture containing all relevant CIs reveals the replicability and stability of the intervals and helps researcher identify the region where the parameter may lie (Wilkinson & APA Task Force on Statistical Inference, 1999). Thompson (in press) wrote, "if we interpret the confidence intervals in our study in the context of the intervals in all related previous studies, the true population parameters will eventually be estimated across studies, even if our prior expectations regarding the parameters are wildly wrong" (p. 21).

Academic journals focus on statistical significance, rather than on documenting and integrating CIs, contributes to a publication bias where only statistically significant results are published, but non-significant results are not, creating an incomplete and biased picture in the literature (Thompson, 2002).

CIs, particularly CIs for effect sizes, also facilitate meta-analytic thinking. Thompson (2002) defined meta-analytic thinking as both the "prospective formulation of study expectations and design by explicitly invoking prior effect sizes" and "the retrospective interpretation of new results,

once they are in hand, via explicit, direct comparison with the prior effect sizes in the related literature" (p. 28). Thinking meta-analytically itself, even absent from other improvements in research practice, Thompson argued, can lead to improved science of discovery (Thompson 2002).

An Evaluation of How Selected Health Education Journals Used Confidence Intervals

To assess how well journals in health education reported and used CIs, an evaluation of articles in two health education journals was conducted. The evaluation aimed to answer two questions: (a) What percentage of articles reported CIs, and (b) what percentage of articles interpreted CIs?

Methods

Two journals of prominent organizations in health education were selected for examination of the use of confidence intervals. The journals are the *American Journal of Health Behavior (AJHB)* and *Health Education & Behavior (HEB)*. The *AJHB* is the official publication of the American Academy of Health Behavior, a research-oriented organization. The mission of the Academy is "to serve as the 'research home' for health behavior scholars whose primary commitment is to excellence in research and the application of research to practice" (American Academy of Health Behavior, 2006). *HEB* is the official publication of the Society for Public Health Education (SOPHE). Established in 1950, SOPHE is "the only professional organization devoted exclusively to public health education and health promotion" (Society for Public Health Education, 2005). It is assumed by the authors that these two journals of prominent organizations in health education reflect the some of the highest quality of research in health education.

Since this article evolved from a paper intended for a graduate level statistics class in April 2006, April 2006 was chosen as the time point to collect articles for evaluation. A total of four issues of journals were considered by the authors to be adequate, considering the fact that this paper served the purpose of a tutorial, rather than a full-blown review. Research articles in the two most recent issues of the *AJHB* and the most recent and the third-most recent issues of *HEB* were included (as of April 2006) in the evaluation. The second-most recent issue of *HEB* was excluded from the evaluation because it was not representative of a typical issue of the journal. This issue was devoted exclusively to a research project—the Trial of Activity for Adolescent Girls, focusing on descriptive statistics (e.g., frequencies; none of the articles included statistical significance testing), and qualitative research (including description of the project, e.g., data collection methods and transferring results to practice). Thirty-three research articles were included in the evaluation.

A total of four issues of journals were considered by the authors to be adequate, considering the fact that this paper served the purpose of a tutorial...

Articles were categorized in methodological design as qualitative research (using focus groups and content analysis as the main method of data collection and analysis) and quantitative research (non-qualitative research). Only quantitative research articles were examined for the reporting and interpretation of CIs. If one or more CIs appeared in an article, the article was recorded as reporting CIs. If an article explained what a CI meant and/or compared if the CIs were different from CIs reported in previous studies, the article was recorded as interpreting CIs. References of the evaluated articles are in an appendix available from the first author. Also available from the first author are four tables documenting the methodological design of each article and whether each quantitative article reported and interpreted CIs. Two of the authors independently coded the articles and were in complete agreements with each other.

Results

Regarding methodological design, the majority of the 33 articles were quantitative. Ninety percent (n=18) of the evaluated *AJHB* articles were quantitative, whereas 84.6% (n=11) of the evaluated *HEB* articles were quantitative. The remaining articles employed qualitative methods.

CIs were reported in approximately half of the evaluated quantitative studies in both journals. However, none of the studies interpreted CIs. Among studies that did not report CIs, one article in *AJHB* (5.6%) and four articles in *HEB* (36.4%) reported standard error intervals, which could be converted to CIs. Thirty-three percent (N=6) of *AJHB* articles and 18.2% (n=2) of *HEB* articles reported neither CIs nor standard error intervals.

Of the twelve articles that reported ORs (odds ratios) using logistic regression, eleven reported CIs for the ORs. Of the four articles reporting the development of a scale or instrument, none reported CIs.

Evaluation Discussion

Although CIs were reported in approximately half of the evaluated quantitative studies, they were not interpreted in any of the studies. The reporting of CIs showed that health education researchers were aware of the importance of CIs. The reporting of CIs could facilitate meta-analyses for future researchers. Nevertheless, the lack of interpretation of CIs indicated that health education researchers might not fully understand the meaning of CIs and consequently could not make use of CIs except for presenting the numbers. Additionally, it was observed that researchers might have reported CIs, only when the statistical packages readily

provided CIs in certain analysis, such as logistical regression. This could be a possible explanation for why 11 of the 12 studies involving ORs reported CIs for ORs. Factor loadings, Chi-square, Cronbach's α, and Pearson's r were the major statistics of four reviewed articles regarding the development of a scale or instrument. It was suspected that authors of these four studies did not report CIs because the statistical packages they used did not readily provide calculations for CIs when the studies' major statistics were computed.

How to Calculate Confidence Intervals Using Statistical Software

One prominent barrier to reporting and interpreting CIs is the fact that widely used statistical software, such as Statistical Package for the Social Sciences (SPSS) and Statistical Analysis Software (SAS), limit CIs to mainly "normal or 'central' t-test statistic distributions" (Smithson, 2001, p. 606), which assume normal distributions of data. For example, output provided by the user-friendly "point and click" options in SPSS does not always give the CIs of the statistics. Therefore, when "noncentral" distributions are needed for computations of CIs for specific statistics, such as Cohen's *d*, η^2, R^2, specific syntax must be used in order for popular statistical software, SPSS and SAS, to provide the CIs. Additionally, according to the University of California Academic Technology Services at University of California (2007):

> In many instances, [users] may find that using syntax is simpler and more convenient than using point-and-click. The use of syntax is also helpful in documenting [the] analysis. It is difficult to take adequate notes on modifications made to the data and the procedures used to do the analyses when using point-and-click. However, documenting what [users] are doing in a syntax file is simple and makes reviewing and/or reconstructing the analysis much easier" (p. 1).

Therefore, this section of the article provides point-and-click, along with syntax, needed to calculate CIs for several statistical analyses.

Smithson (2001) provides SPSS script for computing CIs using "noncentrality parameter for the noncentral F distribution [which] converts that into a confidence interval for multiple (or partial) R^2" (p. 627). Additionally, Duhachek and Iacobucci (2004) and Iacobucci and Duhachek (2003) offer SAS and SPSS syntax for measuring reliability, standard error, and CIs. This provides only two examples of using syntax to compute CIs for specific statistics. Therefore, in addition to Smithson's (2001), Duhachek and Iacobucci's (2004), and Iacobucci and Duhachek's (2003) scripts, Table 1 provides SPSS (Version 14.0) commands and syntax for calculation of CIs for various univariate and multivariate statistical analyses.

Another software utilized to calculate and explore CIs is a graphical software called ESCI (Exploratory Software for Confidence Intervals). ESCI was developed by Geoff Cummings and runs through Microsoft Excel (Cummings & Finch, 2001). This software allows users to (a) explore many CI concepts, (b) calculate and display CIs for personal datasets, (c) "calculate CIs for Cohen's standardized effect size *d*," (d) "explore noncentral *t* distributions and their role in statistical power," (e) "use CIs for simple meta-analysis, using original or [standardized] units," and (f) explore all of the previously mentioned concepts "via vivid interactive graphical simulations" (Exploratory Software for Confidence Intervals, 2006). There are many different ESCI modules available for free download and non-commercial use at http://www.latrobe.edu.au/psy/esci/. These modules were developed with Microsoft Excel 2003.

ZumaStat Statistical Programs provide an additional type of software that is compatible with both Microsoft Excel and versions of 7.0 and higher of SPSS. These programs report CIs for "percentages, correlations, means, standard deviations, variance ratios, differences between correlations, squared correlations, partial correlations, squared partial correlations, squared multiple correlations, group differences in squared multiple correlations, averages of correlations, percent of variance accounted for statistics in ANOVA, single degree of freedom contrasts, odds ratios, relative risks and a wide range of additional statistics" (ZumaStat, 2006, Emphasis on Confidence Intervals section). To read more on ZumaStat programs, please refer to http://www.zumastat.com/Home.htm.

Lastly, an SPSS Tools (Levesque, 2006) internet site is available for use and provides good information on SPSS syntax for calculating CIs for specific statistics. The syntax can be found at http://www.spsstools.net/SampleSyntax.htm#Distributions. These programs, software, and websites provide researchers and practitioners with the appropriate means for calculating CIs, and thus, should help to improve reportage of CIs in future research articles.

How Reporting and Interpretation of CIs Would Enable Research Studies to Yield More Insights

One of the reviewed studies, Vittes and Sorenson (2005), offers an opportunity to show how the reporting and interpretation of CIs would enable the studies to yield more insights on the quality of point estimates and the estimation of the parameter. Vittes and Sorenson reported CIs, but did not interpret the CIs in its own context or in the context of all previous studies. The discussion in the next two sections is based on an actual odds ratio and its CI reported by Vittes and Sorenson.

Reporting CIs Makes a Difference

Vittes and Sorenson (2005) reported CIs, but let us take a moment to see what would happen if we remove one of its

Table 1

Statistical Package for the Social Sciences (SPSS) Commands for Statistical Analyses to Calculate Confidence Intervals (CI) (SPSS, 2006)

Statistical analysis	Possible strategy in SPSS to calculate CIs
GLM Multivariate	Run the GLM Multivariate procedure, under the "analyze" menu in SPSS. Click on Options to provide the 95% CI based on Student's t distribution for the differences between the dependent variables.
GLM Univariate	Utilize the PRINT subcommand, and the PARAMETER keyword with the PRINT subcommand provides CI. For the POSTHOC subcommand in the GLM Univariate analysis, the following keywords provide CI for the Posthoc tests: LSD, SIDAK, BONFERRONI, GH, T2, T3, C, DUNNETT, DUNNETTL, DUNNETTR, TUKEY, SCHEFFE, GT2, GABRIEL. Lastly, when using the CRITERIA subcommand in a GLM Univariate analysis, the keyword ALPHA(n) has two functions. It (a) provides the alpha level under which the power is to be calculated, and (b) identifies the CI level. The value of n should be between 0 and 1 to work properly.
Independent-Samples T Test	Run the Independent-Samples T Test, under the "analyze" menu, then click on Options, which provides 95% CI by default.
Linear Regression	Under the "analyze" menu in SPSS, click on the Linear Regression procedure, and the Save option gives the 95% CI for prediction intervals. Additionally, the Estimates option provides the 95% CI for each regression coefficient or covariance matrix.
Logistic Regression	Under the "analyze" menu in SPSS, click on Logistic Regression, and Options gives the 95% CIs for *exp(B)*. Also, the PRINT subcommand, with the CI(level) keyword provides CI for *exp(B)*. The value identified by (level) must be between 1 and 99.
MANOVA (Multivariate Command)	Use the MANOVA: Multivariate command, and specify a type of analysis in parenthesis after MULTIVARIATE keyword: ROY, PILLAI, WILKS, HOTELLING, BONFER. These keywords provide CI. Additionally, the MULTIVARIATE command on CINTERVAL gives CIs similar to the univariate analysis at the 0.95 level.
Mixed Linear Model	Use the MIXED command in SPSS syntax, and CIN(value) provides CI, and the default value is 95%.
Nonlinear Regression	Utilize the NLR command in SPSS syntax and the BOOTSTRAP subcommand provides CI.
One-Sample T Test	Use the "analyze" menu in SPSS, and under the Compare Means option, click on One-Sample T Test. The Options button provides 95% CI by default.
One-Way ANOVA	Use the "analyze" menu in SPSS, and under the Compare Means option, click on One-Way ANOVA. The Post-Hoc option gives the 95% CI for the mean. Additionally, the STATISTICS command, using SPSS syntax, along with the DESCRIPTIVES subcommand, gives the 95% CI for each dependent variable for each group.
Paired-Samples T Test	Use the "analyze" menu in SPSS, and under the Compare Means option, click on Paired-Samples T Test. The 95% CI for difference in means are displayed by default.
Regression	Utilize the REGRESSION command, and the subcommand, CI, provides 95% CI for the unstandardized regression coefficients. To reset the percent for CI, use CIN[(value)], in which the (value) sets the specified percentage interval utilized with the temporary variable types MCIN (lower and upper bounds for predication intervals of the mean predicated response) and ICIN (lower and upper bounds of prediction intervals for a single observation).
Reliability	Utilize the RELIABILITY Command, and the ICC subcommand, along with the CIN keyword, gives the percent for CI and significance levels of the hypothesis testing. Additionally, the Statistics option gives the 95% CI for the intraclass correlation coefficient (SPSS 14.0 Help Database, 2006).

CIs, leaving only the point estimate—the adjusted odds ratio of 7.52.

This particular adjusted odds ratio indicates that adolescents who own handguns are 7.52 times more likely to have recreational gun use than adolescents who do not own a handgun, while adjusting for all the other variables included in the model. The point estimate may lead the readers to think that handgun ownership is an important predictor of recreational gun use. However, since there is no CI for this odds ratio, we do not know the precision of this odds ratio. By providing the 95% CI of 1.01-55.83, Vittes and Sorenson (2005) enable the readers to estimate by themselves the precision of the odds ratio (although such estimates may be wrong; explanations provided later in the article).

How to Interpret a CI without Comparing it to Previous Studies

Had Vittes and Sorenson (2005) interpreted this CI within its own context (i.e., in the context of this one study, but not in the context of all previous studies), the interpretation could have included the following four points:

1. Ninety-five percent of the CIs constructed with the same method as this study, will contain the true odds ratio for the population.

2. This 95% CI of 1.01-55.83 may or may not contain the true odds ratio for the population.

3. This 95% CI of 1.01-55.83 indicates that adolescents who own handguns are more likely than those who do not own a handgun to have recreational gun use by a factor which can be as low as 1.01 or as high as 55.83, while adjusting for all the other variables included in the model.

4. Without comparing this CI to CIs in previous studies, the CI shows that the 7.52 odds ratio (point estimate) could be imprecise, since the interval appears to be wide. In addition, the lower bound was close to the null hypothesis value of 1.00, indicating handgun ownership may not be an important predictor of recreational gun use. Nevertheless, the precision and replicability of the CI cannot be determined until the CI is compared to all CIs from previous studies.

How to Interpret a CI in the Context of All Previous Studies

Although interpreting a CI in its own context reveals more meanings than not interpreting it at all, the most thoughtful interpretation of CI involves the comparison of the current CI with CIs from all related studies (Thompson, 2006). All relevant CIs, no matter they subsume the null hypothesis parameter value or not, need to be included in the comparison. A better estimate of the parameter can be gained from the comparison. To interpret a CI in the context of all related previous studies, the researcher could (a) construct a graph comprising all CIs for the statistics of interest reported so far, and (b) with the visual assistance of graph, compare the current CI with all related CIs from previous research regarding their width and location.

The following discussion illustrates the interpretation of Vittes and Sorenson's (2005) 95% CI of 1.01-55.83 in the context of all related previous research. Since Vittes and Sorenson did not present any CIs from previous research, CIs used in this discussion are hypothetical and for illustrative purposes only.

Suppose seven studies examined the odds ratio for recreational gun use by gun ownership (v. no gun) in adolescents. All seven studies reported CIs for the odds ratios. CIs for the odds ratio are compiled in Figure 1. The true parameter value will eventually be discovered as researchers continue to compare CIs across studies (Thompson, 2006).

Vittes and Sorenson (2005) could have made the following interpretation of the 95% CI of 1.01-55.83, depending on which interval in the graph represents this CI. If their 95% CI of 1.01-55.83 is interval E, the interval is indeed the widest and not precise. However, since the CI covers a frequently reported area, the researcher might interpret the CI as generally consistent with previous research and might have captured the parameter. If their 95% CI of 1.01-55.83 is interval B, the interval is narrower than most of the CIs from previous studies, and can be interpreted as an improvement in the interval estimate. If their 95% CI of 1.01-55.83 is interval

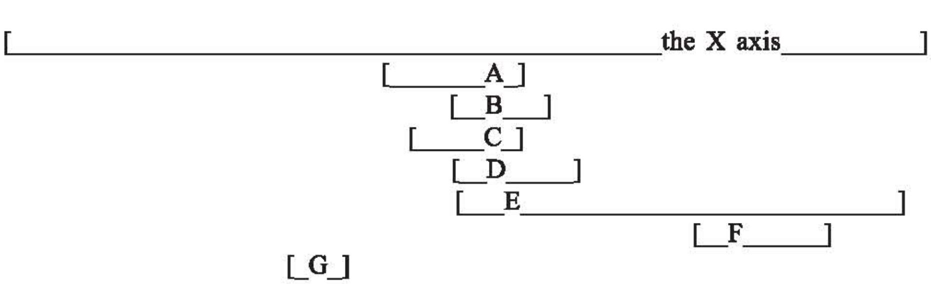

Figure 1. Visual representation of 95% CIs of odds ratio for recreational gun use by gun ownership (v. no gun) in adolescents, reported by all 7 studies.

G, the interval estimate is the narrowest of all the CIs, and may be hastily and happily seen as precise. However, interval G does not cover a frequently reported area. The researcher needs to ponder whether the current CI is accurate and has caught the parameter, or most of the previous CIs are accurate and have contained the parameter. If in fact all the previous CIs contain the parameter, this narrowest CI is inaccurate. The interpretation of a narrow CI as precise demonstrates that simply looking at a CI's width without comparing its location with previous related studies can lead to inaccurate interpretation of the CI. By asking why the current CI is inconsistent with previous CIs, the researchers engage in a critical evaluation of all related CIs in their estimation of the parameter.

Limitations

This article has several limitations. First, the sample size of the evaluated studies (N=33) was too small to generalize to the field of health education. This small study could serve as a pilot study for a full-blown study examining all issues in three to five journals of selected years. Second, causal statements can not be made on the relationship between characteristics and point estimates of studies and whether studies reported CIs.

Conclusion

Making inferences about the population characteristics (parameter) based on knowledge of sample characteristics (statistics) is the goal of inferential statistics (Hinkle et al., 2003). The true parameter value eventually emerges from comparison of CIs for the statistics (Thompson, 2006). Illustrations like Figure 1 assist the comparison of CIs across studies and demonstrate meta-analytic thinking. Schimdt (1996) argues, "Unlike traditional methods based on significance tests, meta-analysis leads to correct conclusions and hence leads to cumulative knowledge" (p. 119).

CIs are the building blocks of the meta-analytic thinking. When CIs for point estimates are not reported, the building blocks for meta-analytic thinking are missing. Without the building blocks, a figure revealing the location of the true parameter cannot be built. When CIs for point estimates are interpreted in the context of a single isolated study, a building block is created and the quality of the building block can be somewhat assessed. We will be able to tell, in some sense, whether a building block is sturdy and usable (narrow) and whether it is flimsy and unusable (wide). However, we cannot know whether a CI is narrow or wide or if it captures the parameter until we compare it with all previous CIs. Without comparing the single CI with all previous CIs, the building block simply lies on the ground and does not contribute to the figure. The full use of the building block is realized only when the CI in the current study is compared to CIs for the same point estimate in all previous related studies. By doing so, the researcher is actively engaged in assessing the quality of his building block, upgrading the quality assessment of previous building blocks, and actually building the figure of meta-analytic thinking. The more researchers add building blocks on the figure, the more the parameter will reveal its location and the more accurate the estimate of the parameter.

The 33 reviewed studies show that health education researchers are beginning to create the building blocks, but are not actively building the figure of meta-analytic thinking. Health education researchers have not fully employed the practice of thinking meta-analytically. However, by utilizing meta-analytic thinking with the assistance of CIs, health education researchers will be able to better estimate the population parameters and use more accurate results to improve people's health.

References

American Academy of Health Behavior. (2006). *Mission Statement*. Retrieved May 8, 2006, from http://www.aahb.org/

American Medical Association. (1998). *Manual of style, a guide for authors and editors* (9th ed). Fairfield, NJ: Author.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Buhi, E. (2005). The insignificance of "significance" tests: Three recommendations for health education researchers. *American Journal of Health Education, 35*, 109-112.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals based on central and noncentral distributions. Educational and Psychological Measurement, 61, 530-572.

Duhachek, A., & Iacobucci, D. (2004). Alpha's standard error (ASE): An accurate and precise confidence interval estimate. *Journal of Applied Psychology, 89*(5), 792-808.

Exploratory software for confidence intervals. (2006). *ESCI Software*. Retrieved June 28, 2006, from http://www.latrobe.edu.au/psy/esci/index.html.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (4th ed.). Boston: Houghton Mifflin.

Iacobucci, D. & Duhachek, A. (2003). Advancing alpha: Measuring reliability with confidence. *Journal of Consumer Psychology, 13*(4), 478-487.

Kline, R. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research.* Washington, DC: American Psychological Association.

Levesque, R. (2006). Raynold's SPSS tools: Syntax. Retrieved June 28, 2006, from http://www.spsstools.net/SampleSyntax.htm#Distributions.

Rivers, D., & Rowell, K. (in press). Encouraging confidence intervals for effect size reporting in health education research. *Eta Sigma Gamma Student Monograph.*

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1,* 115-129.

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurements, 61*(4), 605-632.

Smithson , M. (2003). *Confidence intervals*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-140. Thousand Oaks, CA: Sage.

Society for Public Health Education. (2005). *Mission statement*. Retrieved May 8, 2006, from http://www.sophe.org/content/mission_statement.asp

Statistical Package for the Social Sciences 14.0 (2006). *Help database*. SPSS Inc.: Chicago, IL.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 25-32.

Thompson, B. (2006). *Foundations of behavioral statistics*. New York: Guilford.

Thompson, B. (in press). Research synthesis: Effect sizes. In J. Green, G. Camilli, & P. B. Elmore (Eds.), *Complementary methods for research in education.* Washington. DC: American Educational Research Association.

University of California Academic Technology Services. (2007). Statistical computing seminars beyond point and click: SPSS syntax. Retrieved April 24, 2007, from http://www.ats.ucla.edu/stat/spss/seminars/spss_syntax/default.htm

Vittes, K. A., & Sorenson, S. B. (2005). Recreational gun use by California adolescents. *Health Education & Behavior, 32*(6), 751-766.

Watkins, D. C., Rivers, D., Rowell, K. L., Green, B. L., Rivers, B. (2006). A closer look at effect sizes and their relevance to health education. *American Journal of Health Education, 37*(2), 103-108.

Wiecha, J. M., Lee, V., & Hodgkins, J. (1998). Patterns of smoking, risk factors for smoking, and smoking cessation among Vietnamese men in Massachusetts (United States). *Tobacco Control, 7,* 27-34.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594-604.

ZumaStat. (2006). *ZumaStat statistical programs*. Retrieved June 28, 2006, from http://www.zumastat.com/Home.htm